역전파 알고리듬

히든 레이어가 2개 있는 아래의 신경망에서, 모든 8개의 $\vec{w}$ 각각에 대하여 E의 그래디언트 $\vec{\nabla} E_{\vec{w}}$를 구해보자.

	레이어 1 (히든)	레이어 2 (히든)	레이어 3 (출력)	
	$\vec{w}^1_3, u^1_3, \delta^1_3$		$\vec{w}^o_3, y_3, \delta^o_3$	
x_2	$\vec{w}^1_2, u^1_2, \delta^2_1$	$\vec{w}^2_2, u^2_2, \delta^2_2$	$\vec{w}^o_2, y_2, \delta^o_2$	$E = \frac{1}{2}\sum_{i=1}^{3}(y_i - t_i)^2$
x_1	$\vec{w}^1_1, u^1_1, \delta^1_1$	$\vec{w}^2_1, u^2_1, \delta^2_1$	$\vec{w}^o_1, y_1, \delta^o_1$	

(힌트). ① 다음의 순서대로 δ들을 채워 넣는다: δ^o_1, δ^o_2, δ^o_3, δ^2_1, δ^2_2, δ^1_1, δ^1_2, δ^1_3. 그러면 $\vec{\nabla} E_{\vec{w}^o_i} = \delta^o_i \vec{u}^2$, $\vec{\nabla} E_{\vec{w}^2_i} = \delta^2_i \vec{u}^1$, $\vec{\nabla} E_{\vec{w}^1_i} = \delta^1_i \vec{x}$가 성립한다. 단, 여기서 $\vec{u}^2 = (u^2_1, u^2_2, 1)$, $\vec{u}^1 = (u^1_1, u^1_2, u^1_3, 1)$, $\vec{x} = (x_1, x_2, 1)$를 뜻한다.

② 각 $i = 1, 2, 3$에 대해서 δ^o_i를 구하는 것은 이전과 똑같다. δ^2_i는 δ^o_j들을 계수로 하는 w^o_{ji}들의 선형조합에 $u^2_i(1 - u^2_i)$를 곱한 것이다. δ^1_i는 δ^2_j들을 계수로 하는 w^2_{ji}들의 선형조합에 $u^1_i(1 - u^1_i)$를 곱한 것이다. 이 알고리듬은 행렬곱과 벡터화곱vectorization product을 이용하여 좀 더 간단히 표현할 수 있다.

③ $y = g\big((u_1, u_2, u_3) \cdot (w_1, w_2, w_3)\big)$일 때

$$\frac{\partial y}{\partial w_i} = y(1-y)u_i, \quad \frac{\partial y}{\partial u_i} = y(1-y)w_i, \quad (i = 1, 2, 3)$$ ⊣

Forward

$$\vec{x} = (x_1, x_2, 1)$$
$$u^1_1 = g(\vec{x} \cdot \vec{w}^1_1), \ u^1_2 = g(\vec{x} \cdot \vec{w}^1_2), \ u^1_3 = g(\vec{x} \cdot \vec{w}^1_3), \ \vec{u}^1 = (u^1_1, u^1_2, u^1_3, 1)$$
$$u^2_1 = g(\vec{u}^1 \cdot \vec{w}^2_1), \ u^2_2 = g(\vec{u}^1 \cdot \vec{w}^2_2), \ \vec{u}^2 = (u^2_1, u^2_2, 1)$$
$$y_1 = g(\vec{u}^2 \cdot \vec{w}^o_1), \ y_2 = g(\vec{u}^2 \cdot \vec{w}^o_2), \ y_3 = g(\vec{u}^2 \cdot \vec{w}^o_3)$$
$$E = \frac{1}{2}\big((y_1 - t_1)^2 + (y_2 - t_2)^2 + (y_3 - t_3)^2\big)$$

Backward

$$\frac{\partial E}{\partial w^o_{11}} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial w^o_{11}} = (y_1 - t_1)y_1(1 - y_1)u^2_1$$
$$\vec{\nabla} E_{\vec{w}^o_1} = \delta^o_1 \vec{u}^2, \ \text{where } \delta^o_1 := (y_1 - t_1)y_1(1 - y_1)$$
$$\vec{\nabla} E_{\vec{w}^o_2} = \delta^o_2 \vec{u}^2, \ \text{where } \delta^o_2 := (y_2 - t_2)y_2(1 - y_2)$$
$$\vec{\nabla} E_{\vec{w}^o_3} = \delta^o_3 \vec{u}^2, \ \text{where } \delta^o_3 := (y_3 - t_3)y_3(1 - y_3)$$

$$\frac{\partial E}{\partial w^2_{11}} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial w^2_{11}} + \frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial w^2_{11}} + \frac{\partial E}{\partial y_3}\frac{\partial y_3}{\partial w^2_{11}}$$
$$\frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial w^2_{11}} = (y_1 - t_1)\frac{\partial y_1}{\partial u^2_1} \cdot \frac{\partial u^2_1}{\partial w^2_{11}} = (y_1 - t_1)y_1(1 - y_1)w^o_{11} \cdot u^2_1(1 - u^2_1)u^1_1$$
$$\frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial w^2_{11}} = (y_2 - t_2)\frac{\partial y_2}{\partial u^2_1} \cdot \frac{\partial u^2_1}{\partial w^2_{11}} = (y_2 - t_2)y_2(1 - y_2)w^o_{21} \cdot u^2_1(1 - u^2_1)u^1_1$$

$$\frac{\partial E}{\partial y_3}\frac{\partial y_3}{\partial w_{11}^2} = (y_3 - t_3)\frac{\partial y_3}{\partial u_1^2}\cdot\frac{\partial u_1^2}{\partial w_{11}^2} = (y_3 - t_3)y_3(1-y_3)w_{31}^o \cdot u_1^2(1-u_1^2)u_1^1$$

$$\therefore\ \frac{\partial E}{\partial w_{11}^2} = \delta_1^2 u_1^1,\ \text{where } \delta_1^2 := (\delta_1^o w_{11}^o + \delta_2^o w_{21}^o + \delta_3^o w_{31}^o)u_1^2(1-u_1^2)$$

$$\frac{\partial E}{\partial w_{12}^2} = \delta_1^2 u_2^1,\ \frac{\partial E}{\partial w_{13}^2} = \delta_1^2 u_3^1,\ \frac{\partial E}{\partial w_{14}^2} = \delta_1^2$$

$$\therefore\ \vec{\nabla} E_{\vec{w}_1^2} = \delta_1^2 \vec{u}^1$$

$\vec{\nabla} E_{\vec{w}_2^2}$ 를 구하는 것은 $\frac{\partial E}{\partial w_{21}^2}$ 에서 시작한다. $\frac{\partial E}{\partial w_{11}^2}$ 과의 차이는

$$\frac{\partial y_i}{\partial w_{21}^2} = \frac{\partial y_i}{\partial u_2^2}\cdot\frac{\partial u_2^2}{\partial w_{21}^2} = y_i(1-y_i)w_{i2}^o \cdot u_2^2(1-u_2^2)u_i^1$$

이라는 점이다. 다음의 결과를 얻게 된다.

$$\vec{\nabla} E_{\vec{w}_2^2} = \delta_2^2 \vec{u}^1,\ \text{where } \delta_2^2 := (\delta_1^o w_{12}^o + \delta_2^o w_{22}^o + \delta_3^o w_{32}^o)u_2^2(1-u_2^2)$$

다음 번 레이어에 대한 작업은 조금 더 복잡하지만 이전에 나왔던 계산이 반복되므로 그리 어렵지는 않다.

$$\frac{\partial E}{\partial w_{11}^1} = \frac{\partial E}{\partial u_1^2}\frac{\partial u_1^2}{\partial w_{11}^1} + \frac{\partial E}{\partial u_2^2}\frac{\partial u_2^2}{\partial w_{11}^1}$$

$$\frac{\partial E}{\partial u_1^2} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial u_1^2} + \frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial u_1^2} + \frac{\partial E}{\partial y_3}\frac{\partial y_3}{\partial u_1^2} = \delta_1^o w_{11}^o + \delta_2^o w_{21}^o + \delta_3^o w_{31}^o$$

$$\frac{\partial u_1^2}{\partial w_{11}^1} = \frac{\partial u_1^2}{\partial u_1^1}\cdot\frac{\partial u_1^1}{\partial w_{11}^1} = u_1^2(1-u_1^2)w_{11}^2 \cdot u_1^1(1-u_1^1)x_1$$

$$\therefore\ \frac{\partial E}{\partial u_1^2}\frac{\partial u_1^2}{\partial w_{11}^1} = \delta_1^2 w_{11}^2 u_1^1(1-u_1^1)x_1$$

$$\frac{\partial E}{\partial u_2^2} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial u_2^2} + \frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial u_2^2} + \frac{\partial E}{\partial y_3}\frac{\partial y_3}{\partial u_2^2} = \delta_1^o w_{12}^o + \delta_2^o w_{22}^o + \delta_3^o w_{32}^o$$

$$\frac{\partial u_2^2}{\partial w_{11}^1} = \frac{\partial u_2^2}{\partial u_1^1}\cdot\frac{\partial u_1^1}{\partial w_{11}^1} = u_2^2(1-u_2^2)w_{21}^2 \cdot u_1^1(1-u_1^1)x_1$$

$$\therefore\ \frac{\partial E}{\partial u_1^2}\frac{\partial u_1^2}{\partial w_{11}^1} = \delta_2^2 w_{21}^2 u_1^1(1-u_1^1)x_1 \tag{$*$}$$

$$\therefore\ \frac{\partial E}{\partial w_{11}^1} = (\delta_1^2 w_{11}^2 + \delta_2^2 w_{21}^2)u_1^1(1-u_1^1)x_1 \text{ by } (*1),\ (*2),\ (*3)$$

$\vec{\nabla} E_{\vec{w}_1^1}$ 를 얻기 위해서는 $\frac{\partial E}{\partial w_{12}^1}$ 와 $\frac{\partial E}{\partial w_{13}^1}$ 를 구해야 한다. 이들과 $\frac{\partial E}{\partial w_{11}^1}$ 의 차이는 $(*)$ 의 맨 오른쪽의 x_1 이 각각 x_2 와 x_3 로 바뀐다는 것 밖에 없다. 따라서 다음 결과를 얻는다.

$$\vec{\nabla} E_{\vec{w}_1^1} = \delta_1^1 \vec{x},\ \text{where } \delta_1^1 = (\delta_1^2 w_{11}^2 + \delta_2^2 w_{21}^2)u_1^1(1-u_1^1)$$

남은 2개의 뉴런에 대한 그래디언트는 이런 식으로 꾸준히 계산하면 다음과 같이 얻어진다.

$$\vec{\nabla} E_{\vec{w}_2^1} = \delta_2^1 \vec{x},\ \text{where } \delta_2^1 = (\delta_1^2 w_{12}^2 + \delta_2^2 w_{22}^2)u_2^1(1-u_2^1)$$
$$\vec{\nabla} E_{\vec{w}_3^1} = \delta_3^1 \vec{x},\ \text{where } \delta_3^1 = (\delta_1^2 w_{13}^2 + \delta_2^2 w_{23}^2)u_3^1(1-u_3^1)$$

여기서 주목할 점은 레이어 1의 δ를 계산할 때 직전에 계산해 놓은 레이어 2의 δ만 사용되며 그 이전의 레이어(출력 레이어)는 필요 없다는 것이다. 히든 레이어가 더 많이 있는 일반적인 경우에도 레이어 i의 델타를 계산할 때는 레이어 $i+1$의 값만 사용하며 그 이전의 레이어들에 대한 자료는 필요 없다.

레이어 k의 $\vec{\delta}^k$와 그래디언트 $\vec{\nabla} E_{\vec{w}_i^k}$, $(i = 1, \ldots, n)$를 레이어 $k+1$의 비중벡터 $\vec{w}_i^{k+1}$, $(i = 1, \ldots, m)$과 $\vec{\delta}^{k+1}$으로부터 역전파에 의하여 계산하는 방법을 아래에 정리해 보았다. 레이어 k의 뉴런의 개수는 n, 레이어 $k+1$의 뉴런의 개수는 m이다.

각 $\vec{w}_i^{k+1}$의 길이는 $n+1$이 되어야 한다. 이들에서 마지막 원소를 제거하여 얻은 길이 n의 벡터들을 행벡터로 가지는 행렬 W를 다음과 같이 정의한다.

$$W := \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{pmatrix}$$

δ^{k+1}의 길이는 m이므로 이를 행벡터로 보아 다음의 행렬곱을 얻을 수 있다.

$$\overrightarrow{\delta W} := \vec{\delta}^{k+1} W$$

$\overrightarrow{\delta W}$는 길이 n의 벡터이다. 이제 레이어 k의 $\vec{\delta}^k$는 출력벡터 $\vec{u}^k$를 이용하여 다음과 같이 계산된다.

$$\vec{\delta}^k = \overrightarrow{\delta W} * \vec{u}^k * (1 - \vec{u}^k)$$

위의 식에서 $*$와 $-$는 벡터화 연산vectorized operation이다. 이제 그래디언트 벡터 $\vec{\nabla} E_{\vec{w}_i^k}$들은 레이어 $k-1$의 출력벡터 $\vec{u}^{k-1}$을 이용하여 다음과 같이 계산된다.

$$\vec{\nabla} E_{\vec{w}_i^k} = \vec{\delta}^k * (\vec{u}^{k-1}, 1), \ (i = 1, \ldots, n)$$

$k = 1$일 때는 $\vec{u}^0 := \vec{x}$로 두면 된다.

———————— ∘ ———————— ∘ ————————

backpropagation()의 서브루틴인 append_dg()의 코드에서 벡터들의 리스트로부터 행렬을 만드는 새로운 기법이 사용되었다. do.call() 함수가 핵심인데 이것의 의미는 다음의 코드와 출력을 보면 알 수 있을 것이다.

```
list_of_vectors = list(c(1,2),c(3,4),c(5,6))
print(do.call(rbind, list_of_vectors))
print(do.call(cbind, list_of_vectors))
```

```
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```